DOCUMENT RESUME

ED 220 040 HE 015 399

AUTHOR Whipple, Thomas W.; Muffo, John A.
TITLE Adjusting for Nonresponse Bias: The Case of an Alumni Survey. AIR Forum 1982 Paper.
PUB DATE 19 May 82
NOTE 23p.; Paper presented at the Annual Forum of the Association for Institutional Research (22nd, Denver, CO, May 16-19, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Alumni; Attitude Measures; College Graduates; College Programs; *Graduate Surveys; Higher Education; Institutional Research; Participant Satisfaction; Program Evaluation; *Questionnaires; *Research Problems; *Responses; Response Style (Tests); Sampling; *Statistical Bias; Time; Urban Universities
IDENTIFIERS *AIR Forum; Cleveland State University OH; Mail Questionnaires

ABSTRACT

A method of adjusting for bias due to nonresponse to an alumni survey is demonstrated, based upon analysis of data from both on-time and late respondents. The responses are opinions of 254 alumni regarding how they rate their recently completed program of study within a college of an urban university. As part of the university's curriculum review process, a questionnaire was developed to obtain alumni ratings of their degree program. Twenty-two questionnaire items were selected from the University of Illinois' Program Evaluation Survey (Smock and Hake, 1977), and five demographic questions and a final "would you recommend the program?" question were added. Subjects had received an undergraduate or graduate degree during the last 5 years. It was found that on-time and late respondents differed significantly in their attitudes toward specific aspects of the academic program. Adjustments were made in the average ratings to correct for these differences. In addition, after allowing for random sampling errors, the unadjusted results were contrasted with the adjusted ones. The findings demonstrate that for certain responses, the adjustments were relatively minor, while for other questions the changes were quite significant. In eight cases, the adjusted ratings were significantly less critical of the program than were the unadjusted ratings, based only on the responses of on-time respondents. It is suggested that the increasing reliance on alumni responses to mail questionnaires and their use in the internal review and decision-making process of universities make proper analysis of the data crucial to institutional researchers. The literature on survey nonresponse is also reviewed. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED220040

ADJUSTING FOR NONRESPONSE BIAS:

THE CASE OF AN ALUMNI SURVEY

Thomas W. Whipple
Professor of Marketing
Cleveland State University
1983 East 24th Street
Cleveland, Ohio 44115
(216) 687-4775

John A. Muffo
Director of Institutional Research
Cleveland State University
1983 East 24th Street
Cleveland, Ohio 44115
(216) 687-4700

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Association for

~~Institutional Research~~

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Presented at Association for Institutional Research Annual Forum, Denver, Colorado, May 19, 1982.

HE 015 399

THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

This paper was presented at the Twenty-Second Annual Forum of the Association for Institutional Research held at the Denver Hilton Hotel in Denver, Colorado, May 16-19, 1982. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum papers.

D. R. Coleman, Chairman
Forum Publication
Advisory Committee

Abstract

The increasing involvement of alumni and other external members of the university community in program review has resulted in a reliance on survey instruments to obtain their evaluations. Decision-makers in higher education have found that mail questionnaires are a convenient vehicle for gathering this information. However, the typical low response rates of such surveys have focused researchers' attention on procedures for reducing potential nonresponse bias. This paper demonstrates a method of adjusting for bias due to nonresponse, based upon an analysis of data from both on-time and late respondents. The responses are opinions of 254 alumni regarding how they rate their recently completed program of study within a college of an urban university.

Introduction

Alumni ratings are being used increasingly as a source of evaluative information in universities. According to Wise, Hengstler and Braskamp (1981), the greatest potential utility of alumni evaluations appears to be for purposes of program review. Clark, Hastnett and Baird (1976) argue that recent alumni have a better perspective about the procedures, requirements and contents of a program than do students, and that they tend to be more objective than faculty members. Based on a survey of department chairmen at 134 institutions, Clark (1977) reported that almost 60 percent of the university administrators sampled considered alumni opinions to be very important information in university review and evaluation procedures.

Reliance on input from alumni as part of the program review process necessitates the use of surveys to obtain the opinions sought. A convenient, low cost method which is utilized by many universities is the mail questionnaire. For example, the University of Illinois mails an alumni survey to all degree recipients one year after graduation as a routine practice. As yet, however, the data from these surveys have not been incorporated into the regular review process. Before endorsing the use of alumni ratings, Wise et al (1981) raised several issues dealing with potential sources of measurement error. They addressed these measurement issues to determine what influence, if any, they might have on alumni ratings. Their study demonstrated that alumni ratings do, in fact, provide a relevant and unique source of data which is desirable to include in the assessment of the quality of a department's program.

Their use of questionnaire items from the standard Program

Evaluation Survey (PES), plus the results from their research investigation, allay many expressed concerns about the internal validity of alumni ratings. Nonetheless, they did not address other types of potential errors associated with the use of alumni opinions in the program review process. Of particular concern are external validity and reliability of responses to a mail survey of alumni.

A Research Perspective

External validity and reliability of research results are very common, but also very broad, issues typically raised regarding survey research. Examining the potential sources of these errors allows a researcher either to minimize errors due to specific sources through careful research design, or to adjust for biased results after the fact. Frame and selection sources of external validity errors can be minimized through the use of a sample frame which is representative of the alumni population and a selection procedure which allows each alumnus an equal chance of being chosen to complete the survey instrument. Errors due to nonresponse, however, are not as easily eliminated through research design efforts. When the percentage of nonrespondents is high for a given survey, the prospect for significant systematic bias in the respondent data should be investigated. Also, a low response rate reduces the effective sample size for the study and necessarily decreases the reliability of sample estimates generated from the data. Consequently, preliminary and concurrent efforts to increase the response rate to a survey enhance both the external validity and the reliability of the results. Techniques employed to correct for nonresponse bias, on the other hand, mainly improve the external validity

of the research.

The problem of nonresponse to surveys has haunted survey researchers in all fields for years. Advance notification to sample members and techniques employed regarding the questionnaire itself, the cover letter, the envelope and many other components of the survey may be marginally successful in increasing the response rate. Still, very often, a large percentage of the intended sample has not responded by the deadline. After the fact, the simpliest approach for dealing with potential nonresponse bias is to assume that nonrespondents have the same key characteristics as those responding to the survey. Extrapolating the results of the data analysis to the nonrespondents assumes equality of the two groups, which is simply a matter of blind faith and is not a recommended procedure.

A second tact is to compare known demographics of the nonrespondents to those who respond with the intent of showing that respondents and nonrespondents do not differ significantly based on age, sex, socio-economic status and other demographic variables. Comparing demographics is an improvement over the "blind faith" method, but important attributes being investigated in the survey may be independent of the available demographics.

A more burdensome procedure for measuring nonresponse bias utilizes followup interviews, usually via the telephone, to determine the extent of the differences, if any, between respondents and nonrespondents. The followup procedure is the most desirable of the three methods mentioned, but it is also the most expensive and time consuming. In addition, by altering the medium of questioning,

the followup method raises additional issues of possible interviewer bias and other measurement differences between those responding to mail questionnaires versus telephone calls. Consequently, followup attempts to improve the external validity of a survey may result in the internal validity of the data gathering being questioned.

Purpose

Since each of the above means of dealing with survey nonresponse has major drawbacks, the purpose of this paper is to offer a practical alternative to the methods previously discussed for adjusting for nonresponse bias. An estimate of nonresponse bias and a procedure for correcting the resultant systematic errors in an alumni survey are based on an analysis of on-time and late respondent data.

Reducing and Correcting Nonresponse Bias

The literature of the social sciences and related disciplines has dealt with the problems associated with survey nonresponse in a variety of ways. The greatest amount of attention has been devoted to methods of stimulating responses to mailed questionnaires (Linsky, 1975). Methods to increase the response rate have been classified according to the timing of the efforts and the technique used. Research evidence indicates that advance notification by telephone is effective in increasing response rates and also accelerates the rate of return. However, reminders, especially successive followups, though costly, appear to be a better investment than preliminary notification (Kanuk and Berenson, 1975).

Other efforts to stimulate the response rate include a variety of techniques classified as concurrent methods, since they are incorporated in the major mailing of the questionnaire. In their

review of studies which have focused on the effectiveness of these techniques, Kanuk and Berenson (1975) report that the results regarding questionnaire length, personalization of the mailing, influence of the cover letter, promise of anonymity, questionnaire size, method of reproduction, and color of the questionnaire are inconclusive about the effect on response rate. Their examination of research findings, however, does indicate that official or respected sponsorship, stamped return envelopes, special delivery and air mail postage, and a twenty-five cent incentive are effective in increasing response rates. The results of the studies also showed that deadline dates did not increase the response rate, but did accelerate the speed of questionnaire return. A more recently completed study by Hornik (1981) showed that both response rate and response speed can be enhanced, without causing response bias, by indicating in the cover letter that only a few minutes of the respondent's time was needed to complete the questionnaire.

Another body of knowledge concerning survey nonresponse focuses on how to analyze data already collected so that investigators can account for or correct nonresponse bias. A number of researchers have attempted to identify the salient differences between respondents and nonrespondents so that the degree of bias can be estimated and a correction determined which would make the results of a survey more representative of the sample universe. These efforts have focused on demographic, socio-economic and personality variables.

The only consistent finding is that respondents tend to be better educated than nonrespondents (Kanuk and Berenson, 1975). Wallace (1954) reported virtually no difference between respondents

and nonrespondents in occupation and a number of other socio-economic characteristics, while Robins (1963) discovered higher level occupations among respondents, but no significant differences in social or personality variables. Ognibene (1970), on the other hand, found respondents to be higher in leadership, gregariousness, and reading habits.

Besides educational differences, Donald (1960) suggests that the most promising approach for classifying respondents and nonrespondents may be to assess the interest or involvements of respondents versus nonrespondents in a particular investigation. Her recommendation is based on: (1) evidence that response rates tend to be high when respondents have a special interest or involvement in the subject being studied (Berness, 1960), (2) indications that high levels of interest or involvement are more characteristic of early respondents (Cartwright, 1949), and (2) her own research findings that late respondents are less likely to place a high value on the organization being evaluated (Donald, 1960). Other examinations of early versus late response bias have found no demographic differences, other than in employment and occupational variables (Newman, 1962; Shuttleworth, 1940).

The correlation between speed of response and involvement in the organization has led to efforts to extrapolate trends in responses to estimate nonresponse bias. The basic assumption of such efforts is that respondents who answer later are more like those who do not respond at all than those who answer sooner. If researchers assume that the last wave or combination of the later waves are representative of all the nonrespondents, then they can justify weighting the nonresponses by the late replies. Sometimes it is possible to

establish trends from results of several response waves and then to weight the nonresponse by continuing these trends. However, there is danger in assuming a linear trend. The curve produced may indicate a point beyond which more of the same degree of unfavorable or uninvolved members of the sample will return questionnaires (Donald, 1960). Another problem is that there may be a point at which the direction of the curve changes radically. For example, Baur (1947) discovered that replies from the slowest respondents more closely resembled the earliest respondents than they did the intermediate respondents. If such a U-shaped distribution is not evident, then a simplier mode of analysis than determining a trend is to assume that had nonrespondents given a response, it would have been unlikely to fall at the most favorable end of an attitude scale. Rather than analyzing a sample of nonrespondents and weighting all the nonrespondents according to the results of this analysis, as proposed by Dalenius (1961), it appears reasonable to substitute late respondents for nonrespondents in the weighting formula to derive population estimates.

Research Methodology

The data source for this study consists of responses to a survey of alumni of a single college of a university located in a large, midwestern American city. As part of the University's curriculum review process, a questionnaire was developed to obtain ratings from alumni regarding opinions of their recently completed degree program. To minimize internal validity problems associated with measurement sources of error in the development of the evaluation instrument, twenty-two relevant questionnaire items were selected from the

University of Illinois' Program Evaluation Survey (PES) (Smock and Hake, 1977). In addition, five demographic questions and a final "would you recommend the program?" question were added to complete the instrument.

A list of alumni who had received an undergraduate or graduate degree during the previous five years from the college being evaluated served as the sample frame for the study. The committee conducting the evaluation and university officials believed that less recent alumni would have difficulty recalling certain dimensions of the program under evaluation. Also, curriculum changes six years previously limited the relevant graduation period to five years. The alumni list was culled to eliminate duplicate names, i.e. holders of two degrees. In addition, alumni were instructed to respond to the questions as they related to their most recently completed degree program in the college. A total of 1,749 alumni were selected for the study, each with an equal chance of being chosen from the frame. The maximum sampling error of a sample proportion at the 95% confidence level for a sample of this size is only 2.3 percentage points.

Meticulous questionnaire design, in combination with the use of questions from the PES questionnaire, and exacting sample frame definition and selection procedures reduced the potential for measurement, frame and selection error sources in the alumni survey. Sampling error also would be minimized if the size of the resultant sample approaches the planned sample size. Consequently, the major threat to reducing the validity and reliability of the survey results was a low response rate.

Encorporating the results from previous research regarding concurrent techniques which stimulate the response rate, a cover letter

on university stationary was written by the chairperson of the university's curriculum review committee to authenticate the study. Alumni were advised of the importance of their opinions in the review process, promised that the exercise would take only a "few moments" of their time, guaranteed anonymity, and provided with a stamped return envelope. In addition, since a deadline has been found to accelerate the rate of questionnaire return, it was requested in the cover letter that alumni return the completed questionnaire within ten days. Using similar procedures, including a nonpersonalized cover letter and no followup postcard, Cox, Anderson and Flucher (1974) attained a 13.2% response rate to their evaluative mail questionnaire by the end of a 16-day period.

Findings

By the deadline date, only 120 of the 1,749 questionnaires had been returned to the institutional research office at the university. With the hopes of increasing the response rate from 6.9% to at least the 13.2% achieved by Cox et al (1974), the cutoff date for analyzing the data was postponed one week to allow returned questionnaires to find their way through the university mail system. By weeks end, an additional 31 questionnaires had been received, bringing the total "on-time" response rate to a dismal 8.6%. At that point, time constraints forced the committee to analyze the 151 responses in hand and to report its findings. Average ratings on the twenty-two semantic differential items, as reported for the 151 on-time respondents, are shown in the first column of Table 1.

- - INSERT TABLE 1 HERE - -

Over the next month, 103 "late" questionnaires were received by the

institutional research office. The late response rate of 5.9%, based on the total questionnaires mailed, increased the overall rate of response to the alumni survey to a somewhat more respectable and expected level of 14.5%. However, the late questionnaire data were not analyzed by the curriculum review committee and consequently never were included in the committee report. Although there were no apparent trends in the responses on the later questionnaires, it did appear from a cursory inspection that late respondents were less critical of the program under evaluation.

An analysis of the late responses verified that, in fact, there were no trends in the data. Since it was impossible to establish any trends, the late responses were grouped as representative of opinions of nonrespondents, had they returned their questionnaires. Average ratings on the twenty-two items for the 103 late respondents appear in the second column of Table 1. Comparing these average ratings with those for the on-time respondents revealed a very consistent, systematic bias. For all but the first three evaluative criteria, the late respondents were either equally critical, or in most cases less critical, of the program than the on-time respondents. (For the first three items, a neutral rating, rather than one skewed toward either pole, is probably the most desirous value). This bias is borne out by the additional finding that 87.5% of late respondents, compared to only 80% of on-time respondents, indicated that they would recommend their program of study to interested students.

The results of difference of means tests between on-time and late respondents' scores show in Table 1 that late respondents were significantly less critical than on-time respondents with respect to the following evaluative criteria: (1) the level of the program,

(2) the texts and instructional materials, (3) the quality of instruction, (4) student/faculty contact, (5) the value of the program, (6) student attitudes toward the program, (7) faculty attitudes toward the program, and (8) overall satisfaction with the program. Six of these eight differences are statistically significant at less than the .05 significance level (third column of Table 1), indicating a much less favorable response from on-time respondents. The direction of this response bias is opposite that found by Donald (1960).

Assuming that late respondents are more representative of nonrespondents than are on-time respondents, the average ratings from late respondents were weighted by the large nonrespondent sample size to derive an adjusted average rating for each item. The ratings, adjusted for nonresponse bias, appear in the fourth column of Table 1. The adjusted average ratings were then compared to the 95% confidence interval limits associated with the on-time sample means (shown in the last column of Table 1). It should be noted, that in all eight cases where the on-time and late respondent average ratings are significantly different, the adjusted ratings fall outside the confidence limits. For the other 14 nonsignificant items, the adjusted average ratings are within the confidence limits. In other words, the systematic errors associated with the response bias to eight of the 22 items were greater than the self-compensating, random sampling errors allowed for by a 5% significance level. These findings are consistent with the observation made by Ferber (1948) that: "The problem of response bias must be considered with specific reference to a particular question or characteristic. The presence of bias in one question does not mean <u>a priori</u> that the replies to other questions on the same questionnaire are also biased."

In an attempt to determine if on-time respondents could be distinguished from late respondents, an investigation of demographic and socio-economic differences between the two groups was undertaken. Alumni responses to questions about: (1) the university degree received and the department in which they majored, (2) their occupational status, and (3) how closely their job related to their major course of study were not significantly related to the speed of their responses. As was found in most studies reviewed by Kanuk and Berenson (1975), the data collected in this study provided no basis to conclude that there are discernible demographic or socio-economic differences between on-time and late respondents.

Discussion

The results reported in this paper show that, at least in the case of the one alumni survey described, on-time and late respondents do differ significantly from each other in their attitudes toward specific aspects of the academic program they recently completed. Adjustments are made in the average ratings to correct for these differences; then, allowing for random sampling errors, the unadjusted results are contrasted with the adjusted ones. The findings demonstrate that for certain responses, the adjustments are relatively minor, while for other questions the changes are quite significant. In eight cases, the adjusted ratings are significantly less critical of the program than were the unadjusted ratings, based only on the responses of on-time respondents.

This paper demonstrates to practitioners the why and how of a method which can be used to adjust for nonresponse bias in survey data. The increasing reliance on alumni responses to mail question-

naires and their use in the internal review and decision-making process of universities make proper analysis of the data crucial to institutional researchers. For theoreticians, this study provides added evidence which confirms the usefulness of utilizing strength of attitudes and feelings toward a subject when correcting for non-response bias in survey data.

REFERENCES

Baur, E. J. Response bias in a mail survey, Public opinion quarterly, 1947, 11, 594-600.

Cartwright, D. Some principles of mass persuasion: selected findings of research on the sale of United States war bonds. Human relations, 1949, 2, 264.

Clark, M. J. Program review practices of university departments. Princeton, N. J.: GRE Board Research Report GREB No. 75, July, 1977.

Clark, M. J., Hartnett, R. T., and Baird, L. L. Assessing dimensions of quality in doctoral education. Princeton, N. J.: Educational Testing Service, 1976.

Cox III, E. P., Anderson, Jr., W. T., and Fulcher, D. G. Reappraising mail survey response rates. Journal of marketing research, 1974, 11, 413-417.

Dalenius, T. Treatment of the non-response problem. Journal of advertising research, 1961, 1, 1-7.

Donald, M. N. Implications of non-response for the interpretation of mail questionnaire data. Public opinion quarterly, 1960, 24, 99-114.

Ferber, R. The problem of bias in mail returns: a solution. Public opinion quarterly,1948, 12, 669-676.

Ferness, A. L. A note on stimulating response to questionnaires. American sociological review, 1951, 16, 247.

Hornik, J. Time cue and time perception effect on response to mail surveys. Journal of marketing research, 1981, 18, 243-248.

Kanuk, L. and Berenson, C. Mail surveys and response rates. Journal of marketing research, 1975, 12, 440-453.

Linsky, A. S. Stimulating response to mailed questionnaires: a review. Public opinion quarterly, 1975, 39, 82-101.

Newman, S. W. Differences between early and late respondents to a mailed survey. Journal of advertising research, 1962, 2, 37-39.

Ognibene, P. Traits affecting questionnaire response. Journal of advertising research, 1970, 10, 18-20.

Robins, L. N. The reluctant respondent. Public opinion quarterly, 1963, 27, 276-286.

Shuttleworth, F. K. Sampling errors involved in incomplete returns to mailed questionnaires. Psychological bulletin, 1940, 37, 437.

Smock, H. R. and Hake, H. W. COPE: A systematic approach to the evaluation of academic departments. Paper presented at annual meeting of American Educational Research Associates, New York, April, 1977.

Wise, S. L., Hengstler, D. D., and Braskamp, L. A. Alumni ratings as an indicator of departmental quality. Journal of educational psychology, 1981, 73, 71-77.

TABLE 1

Alumni Program Ratings and Significance Tests

Semantic Differential Items	Respondent Ratings: On-time N=151	Respondent Ratings: Late N=103	Difference of Means Test Significance Levels (α)	Ratings Adjusted for Nonresponse	95% Confidence Interval Limit
1. Framework of program (theoretical/practical)	2.99	2.85	.26	2.87	2.83
2. Amount of structure in curriculum (flexible/rigid)	3.15	3.17	.87	3.16	3.30
3. Orientation of course work (detailed/general)	3.03	2.86	.16	2.88	2.88
4. Variety of course offerings (few/many)	2.60	2.62	.90	2.62	2.77
5. Required courses (integrated/unrelated)	2.43	2.39	.74	2.39	2.27
6. Level of program (easy/difficult)	3.21	3.44	.03[a]	3.42	3.35
7. Program's match to student abilities (challenging/not challenging)	2.48	2.35	.30	2.36	2.33
8. Texts and instructional materials (poor/good)	3.19	3.52	.01[a]	3.49	3.35
9. Quality of instruction (high/low)	2.66	2.45	.09[a]	2.47	2.49
10. Evaluation procedures (unfair/fair)	3.68	3.73	.67	3.72	3.82
11. Enrichment activities (available/unavailable)	3.48	3.49	.99	3.49	3.68
12. Student/Faculty contact (impossible/possible)	3.62	3.83	.09[a]	3.82	3.78
13. Faculty orientation to student needs (concerned/indifferent)	2.53	2.38	.26	2.40	2.36
14. Quality of academic advising (low/high)	2.81	2.80	.99	2.80	2.61
15. Providing credentials for employment (emphasized/ignored)	3.14	2.95	.21	2.96	2.96
16. Quality of vocational counseling (low/high)	2.35	2.42	.66	2.41	2.52
17. Attention to programs (undergraduate/graduate emphasis)	2.41	2.51	.48	2.50	2.57

TABLE 1

Alumni Program Ratings and Significance Tests

Semantic Differential Items	Respondent Ratings: On-time N=151	Respondent Ratings: Late N=103	Difference of Means Test Significance Levels (α)	Ratings Adjusted for Nonresponse	95% Confidence Interval Limit
18. Value of program (worthwhile/useless)	2.08	1.84	.04[a]	1.86	1.93
19. Student attitudes (indifferent/dedicated)	3.30	3.65	.003[a]	3.62	3.45
20. Faculty attitudes (dedicated/indifferent)	2.61	2.27	.01[a]	2.30	2.45
21. Overall satisfaction (very/not at all)	2.46	2.20	.03[a]	2.22	2.30
22. Preparation for professional life (not/very helpful)	3.46	3.64	.24	3.62	3.64

[a]Interpreted as significantly different ratings.